A NOTE ON EXTREMAL DECOMPOSITIONS OF COVARIANCES

ZOLTÁN LÉKA

Abstract. We shall present an elementary approach to extremal decompositions of (quantum) covariance matrices determined by densities. We give a new proof on former results and provide a sharp estimate of the ranks of the densities that appear in the decomposition theorem.

1. Introduction

Let $D \in M_n(\mathbb{C})$ denote an $n \times n$ (complex) density matrix (i.e. $D \geq 0$ and $\mathrm{Tr}\ D = 1$), and let X_i $(1 \leq i \leq k)$ stand for self-adjoint matrices in $M_n(\mathbb{C})$. Then the non-commutative covariance matrix is defined by

$$\mathrm{Var}_D(\mathbf{X})_{ij} := \mathrm{Tr}\ DX_iX_j - (\mathrm{Tr}\ DX_i)(\mathrm{Tr}\ DX_j) \quad 1 \leq i, j \leq k,$$

where $\mathbf{X}$ stands for the tuple $(X_1, \ldots, X_k)$, see [7, p. 13]. We note that there are more general versions of variances and covariance matrices. For instance, in [1], [2] R. Bhatia and C. Davis introduced them by means of completely positive maps and applied the concept for improving non-commutative Schwarz inequalities.

Covariances naturally appear in quantum information theory as well and it seems that there is a recent interest in order to understand their extremal properties [8], [9]. More precisely, in [8] D. Petz and G. Tóth proved that any density matrix D can be written as the convex combination of projections $\{P_l\}$, i.e. $D = \sum_l \lambda_l P_l$, such that

$$\mathrm{Var}_D(X) = \sum_l \lambda_l \mathrm{Var}_{P_l}(X)$$

holds, where X denotes a fixed Hermitian. It is worth it to mention here that quite recently S. Yu pointed out some extremal aspects of the variances which yields a descriptions of the quantum Fisher information in terms of variances (for the details, see [11]).

In this short note we study analogous questions in the multivariable case. Actually, we are interested in the following problem: let us find densities $D_l \in M_n(\mathbb{C})$ such that

$$D = \sum_l \lambda_l D_l \quad \text{and} \quad \mathrm{Var}_D(\mathbf{X}) = \sum_l \lambda_l \mathrm{Var}_{D_l}(\mathbf{X}),$$

where $\sum_l \lambda_l = 1$ and $0 < \lambda_l < 1$. Let us call a density D **extreme with respect to $\mathbf{X} = (X_1, \ldots, X_k)$** if it admits only the trivial decomposition (i.e. $D_l = D$ for every l). It was proved in the cases $k = 1$ and $k = 2$ that the extreme densities are

2000 *Mathematics Subject Classification.* Primary 62J10, 81Q10; Secondary 15B48, 15B57.
Key words and phrases. decomposition, density, covariance, correlation, extreme points.
This study was partially supported by Hungarian NSRF (OTKA) grant no. K104206 .

rank-one projections [6], [8]. Furthermore, the number of projections used, i.e. the length of the decomposition, is polynomial in rank D (see [6]).

The aim of this note is to present a simple approach to the extremal problem above and to look at the question from the theory of extreme correlation matrices (see [3],[4] and [5]). In this context we shall give a new proof to the decomposition theorems appeared in [6], [8], [9] and we present a sharp rank-estimate of the extreme densities.

2. Results and examples

First we collect some basic properties of the covariance matrix $\mathrm{Var}_D(\mathbf{X})$. We note that the matrix does not change by (real) scalar perturbations of the tuple $(X_1, \dots, X_k)$. In fact, an elementary calculation on the entries gives that

$$\mathrm{Var}_D(\mathbf{X}) = \mathrm{Var}_D(X_1 - \lambda_1 I, \dots, X_k - \lambda_k I), \tag{1}$$

where $\lambda_i \in \mathbb{R}$ for every i. Moreover, one can readily check that $\mathrm{Var}_D(\mathbf{X})$ is positive. For the sake of completeness, here is a simple proof.

Lemma 1. $\mathrm{Var}_D(\mathbf{X}) \geq 0$.

Proof. By (1), without loss of generality, one can assume that $\mathrm{Tr}\ DX_i = 0$ holds for every $1 \leq i \leq k$. The density D defines a semi–inner product $\langle A, B\rangle_D := \mathrm{Tr}\ DA^*B$ on $M_n(\mathbb{C})$. Since $\mathrm{Var}_D(\mathbf{X})_{ij} = \langle X_i, X_j\rangle_D$, for any $y = (y_1, \dots, y_k) \in \mathbb{C}^k$, we get that

$$y\mathrm{Var}_D(\mathbf{X})y^* = \langle \sum_i y_i X_i, \sum_i y_i X_i\rangle_D \geq 0$$

and the proof is done. □

Next we show that the covariance is a concave function on the set of the density matrices.

Lemma 2. *Let $D = \sum_l \lambda_l D_l$ be a finite sum of densities $D_l \in M_n(\mathbb{C})$ such that $\sum_l \lambda_l = 1$ and $0 \leq \lambda_l \leq 1$. Then*

$$\mathrm{Var}_D(\mathbf{X}) \geq \sum_l \lambda_l \mathrm{Var}_{D_l}(\mathbf{X}).$$

Proof. Choose $0 < \lambda < 1$. If $D = \lambda D_1 + (1-\lambda)D_2$, a straightforward calculation gives that

$$\mathrm{Var}_D(\mathbf{X}) - (\lambda \mathrm{Var}_{D_1}(\mathbf{X}) + (1-\lambda)\mathrm{Var}_{D_2}(\mathbf{X})) = \lambda(1-\lambda)[x_{ij}]_{1\leq i,j\leq k},$$

where $x_{ij} = \mathrm{Tr}\ (D_1 - D_2)X_i \mathrm{Tr}\ (D_1 - D_2)X_j$. Therefore $[x_{ij}]_{1\leq i,j\leq k} = XX^* \geq 0$ holds with

$$X = \begin{bmatrix} \mathrm{Tr}\ (D_1 - D_2)X_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \mathrm{Tr}\ (D_1 - D_2)X_k & 0 & \dots & 0 \end{bmatrix} \in M_k(\mathbb{C}),$$

and the lemma readily follows. □

The scalar perturbation property $\mathrm{Var}_D(\mathbf{X}) = \mathrm{Var}_D(\mathbf{X} - \lambda)$ guarantees that it is enough to solve the extremal problem when $\mathrm{Tr}\ DX_i = 0$ comes for every $1 \leq i \leq k$. Then the nonlinear part of the covariance vanishes, thus we can simply transform our problem into a geometrical one: let $X_i \in M_n(\mathbb{C})$ $(1 \le i \le k)$ be self-adjoints and define the set

$$\mathcal{D}(\mathbf{X}) := \{D \colon D \in M_n(\mathbb{C}) \text{ is density and} \\ \text{Tr } DX_i = 0 \text{ for every } 1 \le i \le k\}.$$

Clearly, $\mathcal{D}(\mathbf{X})$ is a convex, compact set. From the Krein–Milman theorem, $\mathcal{D}(\mathbf{X})$ is the convex hull of its extreme points. Precisely, these extreme points are the extreme densities we are looking for in the decomposition of $\mathrm{Var}_D(\mathbf{X})$.

Notice that there is no restriction if we assume that $X_1, \dots, X_k$ are linearly independent over $\mathbb{R}$. Hence from here on we shall use this assumption on X_i-s.

When $k \ge 3$, one can see that it is no longer true that the extreme points of $\mathcal{D}(\mathbf{X})$ are rank-one projections. In fact, look at the following simple example in $M_2(\mathbb{C})$ with $k = 3$.

Example 1. Recall that the Pauli matrices are given by

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix} \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Any 2×2 Hermitian Z with $\text{Tr } Z = 1$ can be expressed in the form

$$Z = \frac{1}{2}(I_2 + x\sigma_x + y\sigma_y + z\sigma_z),$$

where x, y and $z \in \mathbb{R}$. Then the points of the Bloch sphere, i.e. $x^2 + y^2 + z^2 = 1$, correspond to the rank-one projections. It is standard that the self-adjoints of trace 1, which are orthogonal to a fixed Z, form an affine 2-dimensional subspace of $\mathbb{R}^3$. Hence one can find X_1, X_2 and X_3 so that the only density D that satisfies $\text{Tr } DX_i = 0$ $(1 \le i \le 3)$ is inside the Bloch ball. Then $\mathcal{D}(\mathbf{X}) = \{D\}$ and D is a density of rank 2.

We shall present a simple characterization of extreme densities or the extreme points of $\mathcal{D}(\mathbf{X})$. We recall that for any positive operators D and C, $D - \varepsilon C$ is positive for some $\varepsilon > 0$ if and only if $\text{ran } C \le \text{ran } D$ holds. Then we can prove

Lemma 3. *The following statements are equivalent:*

(i) *D is an extreme point of $\mathcal{D}(\mathbf{X})$,*

(ii) *if $C \in \mathcal{D}(\mathbf{X})$ such that* $\operatorname{ran} C \le \operatorname{ran} D$ *then $C = D$.*

Proof. Let us assume that $\operatorname{ran} C \le \operatorname{ran} D$ and $D \ne C \in \mathcal{D}(\mathbf{X})$. Then

$$(1 - \varepsilon)\left(\frac{1}{1 - \varepsilon}(D - \varepsilon C)\right) + \varepsilon C = D,$$

where $0 < \varepsilon < 1$, hence D cannot be an extreme point of $\mathcal{D}(\mathbf{X})$.

Conversely, if D is not extreme then $D = \frac{1}{2}D_1 + \frac{1}{2}D_2$ which implies that $\text{ran } D - \frac{1}{2}D_1 \le \text{ran } D$, since $D - \frac{1}{2}D_1$ is positive. □

To produce a description of $\operatorname{ext} \mathcal{D}(\mathbf{X})$ which is more effective for our purposes, we need some basic facts about correlation matrices. We recall that a positive semidefinite matrix is a correlation matrix if its diagonal entries are 1-s. Correlation matrices form a convex, compact set in $M_n(\mathbb{C})$. Its extreme points, or extreme correlation matrices, were described by several authors, see e.g. [4], [5]. It is well-known that an $n \times n$ extreme correlation matrix has rank at most $\sqrt{n}$ (see e.g. [3]).

Later we shall present an estimate of the rank of extreme densities matrices (with respect to tuples).

The perturbation method used by C.-K. Li and B.-S. Tam is relevant for us. Let us say that a nonzero Hermitian $S \in M_n(\mathbb{C})$ is a **perturbation** of D if there exists an $\varepsilon > 0$ such that $D \pm \varepsilon S$ are density matrices as well. Then D is an extreme density with respect to $X_1, \dots, X_k$ if and only if there does not exist perturbation S of D such that $\mathrm{Tr}\ S = 0$ and $\mathrm{Tr}\ SX_i = 0$ for every $1 \le i \le k$. In fact, if D is not extreme, one can find D_1 and D_2 densities such that $D = \frac{1}{2}D_1 + \frac{1}{2}D_2$ and $\mathrm{Tr}\ D_j X_i = 0$. It follows that $S = D_1 - D_2$ is a perturbation of D. The converse statement is trivial.

From here on let $H_n(\mathbb{C})$ denote the real Hilbert space of $n \times n$ complex Hermitian matrices with the usual inner product $\langle A, B\rangle = \mathrm{Tr}\ AB$. One can easily conclude that an extreme density D (with respect to $\mathbf{X}$) must be singular if $n^2 > k+1$. Actually, the last inequality guarantees the existence of a Hermitian perturbation S which satisfies the orthogonality constraints; i.e. S is orthogonal to X_i-s and I. Moreover, the continuity of the spectra here gives that any small perturbation $D \pm \varepsilon S$ is positive if D is invertible.

Let $\sigma(A)$ denote the spectrum of any $A \in M_n(\mathbb{C})$. Suppose that the matrix D is of rank r. Then there does exist an $Y \in M_{n\times r}(\mathbb{C})$ and $R \in H_r(\mathbb{C})$ such that $D = YRY^*$. Now one can prove the following lemma which is analogous to [5, Theorem 1. (a)].

Lemma 4. *Let $D = YRY^* \in \mathcal{D}(\mathbf{X})$ be a density of rank r. Then S is a perturbation of D if and only if $\mathrm{Tr}\, S = 0$ and $S = YQY^*$ where $Q \in H_r(\mathbb{C})$.*

Proof. First, assume that $S = YQY^*$. Then S is nonzero if and only if $Q \neq 0$. Indeed, we have rank $S =$ rank Q because Y has full column rank r. Since $D = YRY^*$ is positive, we obtain that R is positive and invertible. From $0 \notin \sigma(R)$, there does exist an $\varepsilon > 0$, such that $D \pm \varepsilon S = Y(R \pm \varepsilon Q)Y^*$ are positive. Obviously, we get that S is a perturbation.

Conversely, let us assume that S is perturbation of D. Clearly, $\mathrm{Tr}\ S = 0$ must hold. Expand Y with a matrix $Z \in M_{n\times(n-r)}(\mathbb{C})$ such that $V = (Y|Z)$ is invertible and $V(R \oplus 0_{n-r})V^* = D$ hold. Next, let us write $V^{-1}S(V^*)^{-1}$ into blocks that corresponds to the block form of $R \oplus 0_{n-r}$. Since $V^{-1}(D \pm \varepsilon S)(V^{-1})^*$ are positive for some $\varepsilon > 0$, it follows that $S = V(Q \oplus 0_{n-r})V^*$ must hold for some $Q \in H_r(\mathbb{C})$. □

After this lemma here is our main result which reflects some similarity with the characterization theorem of extreme correlations, see [5, Theorem 1].

Theorem 1. *Let $X_i \in H_n(\mathbb{C})$, $1 \le i \le k$, and $D = YRY^* \in \mathcal{D}(\mathbf{X})$ be a density of rank r, where $Y \in M_{n\times r}(\mathbb{C})$. The followings are equivalent:*

(i) *D is an extreme point of $\mathcal{D}(\mathbf{X})$,*
(ii) $\mathrm{span}\,\{Y^*X_1Y, \dots, Y^*X_kY, Y^*Y\} = H_r(\mathbb{C})$,
(iii) $\{DX_1D, \dots, DX_kD, D^2\}$ *has (real) rank r^2.*

Moreover, if $D = YY^*$ then the above statements are equivalent to

(iv) *$r^{-1}I_r$ is an extreme density with respect to $Y^*\mathbf{X}Y$; that is,*

$$\mathcal{D}(Y^*\mathbf{X}Y) = \{r^{-1}I_r\}.$$

Proof. (i) $\Leftrightarrow$ (ii) From Lemma 4, D is extreme if and only if there does not exist $0 \neq YQY^*$ such that Tr $YQY^*Y_i = \text{Tr } Q(Y^*X_iY) = 0$ and Tr $YQY^* = \text{Tr } Q(Y^*Y) = 0$. We notice that $Q = 0$ if and only if the linear span of $Y^*X_1Y, \ldots, Y^*X_kY$ and Y^*Y is the full space $H_r(\mathbb{C})$.
(iii) $\Leftrightarrow$ (ii) Let us choose the decomposition $D = YY^*$; that is, $R = I_r$. Note that the self-adjoint $Y^*Y \in M_r(\mathbb{C})$ is invertible. In fact, $\sigma(YY^*) \cup \{0\} = \sigma(Y^*Y) \cup \{0\}$ holds, thus $\sigma(Y^*Y)$ equals to the set of positive eigenvalues of D (with multiplicities). This implies that $\sum_{i=0}^k \alpha_i Y^*X_iY = 0$ if and only if $\sum_{i=0}^k \alpha_i YY^*X_iYY^* = 0$ ($\alpha_i \in \mathbb{R}$, $X_0 = I_n$), so the systems $\{Y^*X_1Y, \ldots, Y^*X_kY, Y^*Y\}$ and $\{DX_1D, \ldots, DX_kD, D^2\}$ have the same rank.
(i) $\Rightarrow$ (iv) Since D is an extreme point, we get from (ii) that $\{Y^*X_1Y, \ldots, Y^*X_kY\}$ has rank at least $r^2 - 1$. However, I_r is not in the linear span of the above system because it is orthogonal to every matrix Y^*X_iY. Adjusting $r^{-1}I_r$ to $Y^*\mathbf{X}Y$, we get a full rank system of $H_r(\mathbb{C})$. Hence by (iii) we conclude that $r^{-1}I_r$ is an extreme point of $\mathcal{D}(Y^*\mathbf{X}Y)$.
(iv) $\Rightarrow$ (i) If $r^{-1}I_r$ is an extreme point, it has no perturbation S which is orthogonal to every Y^*X_iY. Thus it follows that $I_r, Y^*X_1Y, \ldots, Y^*X_kY$ must span $H_r(\mathbb{C})$; that is, $\mathcal{D}(Y^*\mathbf{X}Y) = \{r^{-1}I_r\}$. Note that $Y^*Y, Y^*X_1Y, \ldots, Y^*X_kY$ span $H_r(\mathbb{C})$ as well becase Tr $Y^*Y = \text{Tr } D = 1$ and Y^*X_iY-s are traceless. Thus (ii) implies that D is an extreme point. $\square$

The theorem gives a straightforward estimate of the rank of extreme densities.

Corollary 1. *Let $D \in M_n(\mathbb{C})$ be an extreme density with respect to $X_1, \ldots, X_k \in H_n(\mathbb{C})$. Then*
$$\text{rank } D \leq \sqrt{k+1}.$$

The Krein–Milman theorem implies that $\text{Var}_D(\mathbf{X})$ can be written as the convex sum of covariances determined by densities of rank at most $\sqrt{k+1}$. Moreover, one can easily deduce the following result which first appeared in [6], [9] and [8, Theorem].

Corollary 2. *Let $D \in M_n(\mathbb{C})$ denote a density matrix. In the case of $k = 1$ and $k = 2$, there exist projections $P_1, \ldots, P_m$ such that*
$$D = \sum_{l=1}^m \lambda_l P_l \quad \textit{and} \quad \text{Var}_D(\mathbf{X}) = \sum_{l=1}^m \lambda_l \text{Var}_{P_l}(\mathbf{X})$$
hold, where $\sum_{l=1}^m \lambda_l = 1$ and $0 \leq \lambda_l \leq 1$.

In the case of $k \geq 3$, one might expect that the covariance matrix still can be decomposed by means of projections if n is large enough. However, this is not necessarily true. The next example shows that the estimate of Corollary 1 is sharp if n is large enough.

Example 2. Let $n = \lfloor \sqrt{k+1} \rfloor$. The special unitary group $SU(n)$ has dimension $n^2 - 1$, so let λ_i ($1 \leq i \leq n^2 - 1$) denote a collection of its traceless, Hermitian infinitesimal generators. One can also assume that Tr $\lambda_i \lambda_j = 0$ holds for every $i \neq j$ (for the generalized Gell–Mann matrices, see e.g. [10]). Then the matrices $\{I_n, \lambda_1, \ldots, \lambda_{n^2-1}\}$ span the real vector space $H_n(\mathbb{C})$. Thus it follows that
$$\mathcal{D}(\lambda_1, \ldots, \lambda_{n^2-1}) = \left\{ \frac{I_n}{n} \right\}$$

is a singleton, hence $(1/n)I_n$ is an extreme density of rank n. If $n^2 < k+1$, let us choose arbitrary $\lambda_{n^2}, \ldots, \lambda_k \in M_m(\mathbb{C})$ Hermitians which are linearly independent where m is large enough. From Theorem 1 (iii), $(1/n)I_n \oplus 0_m$ remains extremal with respect to $\lambda = (\lambda_1 \oplus 0_m, \ldots, \lambda_{n^2-1} \oplus 0_m, 0_n \oplus \lambda_{n^2}, \ldots, 0_n \oplus \lambda_k)$, hence $\mathrm{Var}_{(1/n)I_n \oplus 0_m}(\lambda)$ is not decomposable.

Applying direct sums as above, for every large n one can construct $n \times n$ extreme densities of arbitrary rank between 1 and $\sqrt{k+1}$.

The method we used is very similar to that of describing extreme correlations. However, the next example shows that $\mathrm{Var}_D(\mathbf{X})$ is not necessarily extreme even if it is a correlation matrix and D is an extreme density (with respect to some tuple).

Example 3. Let D be the projection $\mathrm{diag}(1, 0, \ldots, 0) \in \mathbb{R}^{n+1}$. We define the Hermitians in $H_{n+1}(\mathbb{C})$

$$X_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \oplus 0_{n-1}, \; X_2 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \oplus 0_{n-2}, \; \ldots \,,$$

$$X_n := \begin{bmatrix} 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & 0 \\ 0 & \vdots & \vdots & \vdots \\ 1 & 0 & \ldots & 0 \end{bmatrix}.$$

Then a simple calculation gives that $\mathrm{Var}_D(\mathbf{X}) = I_n$ which is obviously not an extreme correlation matrix.

Finally, for the converse, we give an example that $\mathrm{Var}_D(\mathbf{X})$ can be an extreme correlation matrix while D is not necessarily extremal (with respect to $\mathbf{X}$).

Example 4. Consider $D = (1/n)I_n \oplus 0_n \in H_{2n}(\mathbb{C})$, $n > 2$. Let us choose reals $x_1, \ldots, x_n$ such that $\sum_{i=1}^n x_i = 0$ and $\sum_{i=1}^n n x_i^2 = 1$ hold. For any $\tilde{X}_i \in H_n(\mathbb{C})$, $1 \le i \le n$, we set

$$X_i = \mathrm{diag}(x_1, \ldots, x_n) \oplus \tilde{X}_i \in H_{2n}(\mathbb{C}) \qquad 1 \le i \le n.$$

Then we get that $\mathrm{Var}_D(\mathbf{X})$ is the $n \times n$ matrix which consists only 1-s; that is, it is a rank-one extreme correlation matrix. From Corollary 1, D cannot be extreme with respect to $\mathbf{X}$.

References

[1] R. Bhatia, Positive Definite Matrices, Princeton University Press, Oxford, 2007.

[2] R. Bhatia and C. Davis, More operator versions of the Schwarz inequality, Commun. Math. Phys., **215** (2000), 239–244.

[3] J.P.R. Christensen and J. Vesterstrøm, A note on extreme positive definite matrices, Math. Ann., **244** (1979), 65–68.

[4] R. Grone, S. Pierce and W. Watkins, Extremal correlation matrices, Linear Alg. and Its Appl., **134** (1990), 63–70.

[5] C-K. Li and B-S. Tam, A note on extreme correlation matrices, SIAM J. Matrix Anal. Appl., **15** (1994), 903–908.

[6] Z. Léka and D. Petz, Some decompositions of matrix variances, Probab. Math. Stat., **33** (2013), 191–199.

[7] K.R. Parthasarathy, An Introduction to Quantum Stochastic Calculus, Birkhäuser Verlag, Basel, 1992.

[8] D. Petz and G. Tóth, Matrix variances with projections, Acta Sci. Math. (Szeged), **78** (2012), 683–688.
[9] D. Petz and G. Tóth, Extremal properties of the variance and the quantum Fisher information, Phys. Review A, **87** (2013), 032324
[10] H–J. Sommers and K. Zyckowski, Hilbert–Schmidt volume of the set of mixed quantum states, J. Phys. A, **36** (2003), 10115–10130.
[11] S. Yu, Quantum Fisher information as the convex roof of variance, preprint, arXiv:1302.5311.

Alfréd Rényi Institute of Mathematics, 1053 Budapest, Reáltanoda u. 13-15, Hungary
E-mail address: leka.zoltan@renyi.mta.hu